Voice Activated GUI – The Next User Interface

Rosalind Rogoff
My Training Dept., Inc.
9913 Mangos Drive
San Ramon, CA 94583
info@mytrainingdept.com

Voice Activation is the user interface of the future. Not only will you be able to tell your computer what to do, you will be able to activate multiple functions in your car, television, or microwave oven by voice commands. This session considers how voice activation could impact your life and your job, and how the field of technical communications will change because of it. Rosalind Rogoff will demonstrate voice commands and dictation using L&H VoiceXpress and the Conversay Voice Surfer. This session answers the following questions:

- *What is Voice Activation?*
- *What can I do with it?*
- *What hardware and software do I need?*
- *What are the differences between VUI and VA-GUI?*
- *What might the future hold?*

Keywords: Voice Activation, User Interface, Speech Recognition, VoiceXML

WHAT IS VOICE ACTIVATION

Voice activation is using your voice to control your PC or other aspects of your environment. Dictation software has been around for over five years. It is used for writing letters and email, but can also be used to open documents, save documents, dial the phone, close or exit a program, and most other standard commands in Microsoft Word or other common MS applications.

Some of the more sophisticated dictation software, such as IBM's Via Voice Pro, or Dragon Naturally Speaking Professional, can be used with voice macros to trigger a series of functions by voice command (Newman, 2000). These applications are used by people who can't use a keyboard or mouse. They are becoming more popular with people who have suffered, or want to prevent, Carpal Tunnel Syndrome. Professional

level dictation applications are available with special vocabularies for Doctors and Lawyers, and even teenagers. You can add your own list of words to these applications and create your own lexicon.

Via Voice comes with a Voice Mouse option, which is literally a voice controlled mouse. A little mouse icon can be positioned on your PC desktop by voice, and then told to click on an icon or menu to perform that function. Realize software uses just the command portion of voice activation to control the desktop or open applications, without the dictation option.

Voice activation is available now in some new luxury cars. It's easy to see why automobiles are an appealing venue for voice activation. No more fumbling for the right switch on the radio or dashboard. Simply say "headlights on," and the headlights go on. These options are pricey right now, but will filter down to lower priced vehicles in the next few years.

HOW CAN YOU USE VOICE ACTIVATION?

Dictation is still the main function of popular voice activation software. These applications are speaker dependent, and require many hours reading text into your computer for accurate recognition.

The Conversay voice surfer is a plug-in for a web browser that controls all web navigation by voice. Command software, like Realize and Conversay, are speaker independent, and do not require extensive personalized training. When you use these applications, you can ask for a list of commands, usually by saying "What can I say?" Since this list is fairly short, there isn't as much chance of the computer misrecognizing what you are saying.

Right now voice recognition companies are heavily marketing embedded applications. Cell phones that dial a number when you say "phone Mom," are a good example. Expect voice activated toys to be a hot item this Christmas.

Biometrics is another application of voice activation that is on the verge of taking off. Some voice recognition programs include speaker verification that can be used as part of voice activated security systems. If the sample of users is small, and the vocal differences are significant, such as high pitch vs. low, or male vs. female, voice security can be effective. However, these systems are not foolproof yet, and shouldn't be used where there might be genuine security risks (Rodman, 1999).

WHAT DO YOU NEED TO GET STARTED?

Hardware	CPU	Pentium III/450 MHz or faster
	RAM	128 MB
	Hard-Drive	8+GB hard disk
	Sound	Creative Labs Soundblaster or compatible
	Graphics	16 MB Graphics card
	Microphone	Good quality noise canceling mike
Software	Dictation	IBM ViaVoice, L&H VoiceXpress, Dragon Naturally Speaking
	Translation	L&H Power Translator Pro
	Command	Realize, Conversay, ViaVoice Pro
	Web Surfing	Conversay Voice Surfer

This is not a complete list of voice activation software, but it will get you started. You can purchase dictation software online or from an office supply store. You can download Conversay Voice Browser or Realize Voice Lite from their web sites. Boxed software comes with a moderately good headset microphone, usually an Andrea 61. You can get a better one from an electronics supply store.

You will need to spend several hours "training" the software, by reading prepared texts into your computer. When you first start dictating, you will probably have to correct a lot of errors and do more "training" sessions. If you are testing command or web surfing software, you should be able to control desktop functions, create voice macros, or surf the web, with hardly any software training at all.

VOICE USER INTERFACE VS. VA-GUI

Voice User Interface is the telephone application of voice activation. Right now this is the hottest application of voice recognition. There are millions of cell phones in the world, and people are using them everywhere, creating a huge market for easy access to information.

Cell phone services offer "virtual assistants," which provide a range of information. You can call and ask for the weather, stock prices, sports scores, email, or directions to the nearest restaurant. Some of the early systems simply updated old IVR (interactive voice response) systems with vocalizations of keypad entries such as, "say one for sales, say two for service." Motorola predicted the market for virtual assistants would be $3.4 billion dollars by 2003 (Janal, 1999).

Nuance and SpeechWorks offer voice web servers, which extend the virtual assistant concept to web content and web-like portals for navigating to a wider range of information. Voice portals are used for customer service. Banks and airlines want the interaction with their customers to sound natural. This has spawned a whole new industry, writing Voice User Interface scripts (Balentine, 1999). These are usually written in VoiceXML, which is an extension of standard XML. The challenge of VoiceXML isn't just the programming, but anticipating what the user might say, and coming up with a list of synonyms for every possible utterance. These are called grammars (Edgar, 2001), and are used to make the dialog with the voice on the other end of the phone (usually prerecorded, but possibly text-to-speech), seem like talking to a real person.

With VA-GUI, you know you are talking to your computer, or refrigerator, or car. You don't have to pretend that this is anything more natural than clicking a mouse or flipping a switch with your voice. Switches have labels, and hypertext links can be verbal, so it is easy to know what to say. In addition VA-GUI can display a visual list of commands. So you don't need special programming or extensive grammars to make the conversation seem natural.

Conversay promotes itself as "conversational computing." You can add audio prompts and responses to web pages with the Conversay Voice Surfer Tool Kit, but I don't think most people really want to have conversations with their computers. We aren't chatting with HAL of "2001," or cute little Haley Joel Osment of "AI." We are talking to machines, and we don't have to be tactful, courteous or afraid.

DEMOS OF VOICE ACTIVATED GUI (AS TIME ALLOWS)

- L&H VoiceXpress – Commands and dictation.
- Conversay - Voice Browsing

WHERE DO WE GO FROM HERE?

The dot.com revolution seemed to come on like a great wave that hit the shore and died. The wave of Voice Activation that is coming won't die off. Technology already exists to translate other languages, and read or reproduce what someone says with text-to-speech (TTS). Employees will be able to ask for instructions on the job, and in whatever language they speak. As voice security improves, we could even vote by voice, which would probably be more reliable than the butterfly ballot. The job of the technical communicators will have the added dimensions of working with, and developing for, all these voice activated systems.

REFERENCES

Balentine, Bruce , David P. Morgan. How to Build a Speech Recognition Application. 141-164, 1999.

Edgar, Bob. The VoiceXML Handbook. 181-196. 2001.

Janal, Daniel S. Business Speak. 183-192. 1999.

Newman, Dan. Talk to Your Computer. 115-123. 2000

Rodman, Robert D. Computer Speech Technology. 220-249. *1999.*

ABOUT THE AUTHOR

Rosalind Rogoff has a Ph.D. in Instructional Design and over 20 years of experience as a Systems Trainer and Technical Writer. She was President of the East Bay Chapter of the STC for 2000-01. Her book, The Training Wheel: A Simple Model for Instructional Design, was published by John Wiley & Sons in 1987. Rosalind is Documentation Manager for Relevant Business Systems in San Ramon. She recently formed My Training Dept., Inc., www.mytrainingdept.com, to develop Hands-Free Help™ using Conversay voice recognition technology. Rosalind is a certified Conversay Developer, Authorized Training Provider, and Gold Member of the Conversay Developers Network.